DOCUMENT RESUME

ED 212 657 TM 820 049

AUTHOR	Polin, L.; Baker, E.L.
TITLE	Exploration of "Fuzzy Set" Procedures for Analysis of Item-Specification Congruence.
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	Nov 78
GRANT	OB-NIE-G-78-0213
NOTE	9p.
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Rating Scales; *Test Construction; *Test Items; Test Selection; *Test Validity
IDENTIFIERS	Content Validity; Domain Specifications; *Fuzzy Set Theory; *Item Rating Scale

ABSTRACT

A neglected element in designing tests is that of publicness, that is, the extent to which test specifications are understandable and usable by all interested parties. Issues related to content validity, such as test bias and instructional sensitivity, become accessible to these parties once content validity and design have been adequately described. Applying mathematical and set theory concepts, this paper explores procedures for determining content validity based upon "judged" degree of match between domain specification and item. The notion of degree of "belongingness" of an item to its domain and accompanying "fuzzy set" rating scale were applied to rating the match between given test items and test specifications. Since the form of specification did not provide sufficient guidance to raters for consistent and reliable judgments, an Item Rating Scale was created to distinguish important item dimensions in achievement tests and to thus provide structure for raters' judgment of congruence. The instrument is viewed as a training intervention for those responsible for test selection, item selection from banks for the purpose of assessing objectives, and for test item construction and review. Plans for review and completion of this study and for presentation of results are indicated. (Author/AL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED212657

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

DELIVERABLE: NOVEMBER 1978

STUDIES IN MEASUREMENT & METHODOLOGY

Work Unit 1:
Design and Use of Tests

Eva L. Baker
Project Director

EXPLORATION OF "FUZZY SET" PROCEDURES FOR ANALYSIS OF ITEM-SPECIFICATION CONGRUENCE

L. Polin & E. L. Baker

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
G.S. Gray
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Grant Number
OB-NIEG-78-0213

Center for the Study of Evaluation
Graduate School of Education
University of California, Los Angeles

TM 820 015

The work upon which this publication is based was performed pursuant to a grant with the National Institute of Education, Department of Health, Education and Welfare. Points of view or opinions stated do not necessarily represent official NIE position or policy.

INTRODUCTION

One of the most neglected elements in designing tests is that of publicness; that is, the extent to which test specifications are understandable and usable by all interested parties, including teachers, parents, taxpayers and even students themselves. An entire set of content validity related issues, such as test bias and instructional sensitivity become accessible to these parties once test content and design have been adequately described. Domain-referenced testing, in particular, is based upon the assumption that by limiting and defining a class of behaviors, skills, information (i.e., a domain), a set of rules may be created for generating sets of test items. The degree to which these items accurately reflect the content of the domain affects the power of such tests to judge student test performance as an accurate indicator of student competence in that domain. Content, or descriptive, validity is concerned with the extent to which test items comprise a representative sampling from the domain.

At present there are no workable methodologies for determining content validity, especially under the criterion of publicness as described above. Guion (1977) in psychological measurement and Popham (1975, 1978) in educational evaluation, describe the methodology for determining content validity using the term "judgment," referring to the individual's thoughtful but subjective decision on the quality of fit between domain and items. Other methodologies attempting to systematize and objectify the judgment process have been explored and found lacking. In an earlier CSE project, item-objective congruence methods of sorting, matching and even

generating objectives from given items were tried and found to be unreliable (Walker, Trithart and Polin, 1976).

Method

The basic concept of making validity determinations based upon "judged" degree of match between each item and the domain specification remains appealing largely because it is comprehensible to a wide range of persons interested in testing, test development and selection. In our approach to the content validity judgment issue we have borrowed from the field of mathematics and set theory. Based upon work by Kaufman (1975) and Rhodes and Tilson (in Arbib, 1968), the notion of estimating "belongingness" of an item to its domain or set, has guided our work. This idea of "fuzzy set" theory allows the rating of the degree of match between specification and item, specifying the strength of the relationship on a 0 to 1 continuum. Conceptually, the notion is similar to probability; that is, with what probability does one think an item belongs to a particular set.

In our first study, this notion of degree of belongingness and the accompanying fuzzy set rating scale were applied to rating the match between given test items and test specifications. It became clear that the form of specification did not provide sufficient guidance to raters for consistent and reliable judgments. To remedy this problem and continue research into the application of fuzzy set judgments to content validity, we developed an Item Rating Scale (I.R.S.). Paralleling work by Carroll (1976) in the specification of intelligence tests, our Item Rating Scale was created to distinguish important item dimensions in achievement tests and to thus provide structure for raters' judgments of congruence. The IRS identifies

four major categories of item features: Subject Matter, Cognitive Processing Complexity, Syntactic/Semantic Complexity, and Format.

The IRS has been subjected to internal review and revision. Additionally, the revised instrument has been used with graduate students in the UCLA doctoral course on instructional product development. Originally we had planned to re-train raters and rerun the initial study using items for which student performance data exists. At this point in time, however, we find ourselves with the unique opportunity of presenting our work for review and field testing in the 1978-79 Regional Testing Conferences sponsored by NIE and managed by CSE. The advantage of interfacing with the testing conferences is that we will have access both to leaders in domain-referenced measurement and to practitioners responsible for making test design and selection decisions. We still plan to train raters to make fuzzy set judgments of item-specification congruence considering the dimensions outlined on the IRS. We see the instrument as useful as a training intervention for those charged either with the selection of tests, the selection of items from pools or banks for the purpose of assessing given objectives, and for those whose task is test item construction and review.

Specifically, within the next period we will:

1. refine the IRS for use by teachers;
2. explore and construct necessary descriptive material to explain the intent of the instrument;
3. develop sample specifications and item sets which are designed to exhibit or exclude certain specified item attributes;
4. conduct a study to determine the minimal feasible

training required to use in this artificial setting;

5. use the instrument with "real" specifications and item sets in a study. We expect to use items for which performance data exist so that the critical features of the decisions of test item-specification congruence can be studied.

Accordingly, our efforts are currently directed toward acquisition and compilation of field test material and formal review guidelines. As the field test and review data from each conference are analyzed, we will have the opportunity to make further modifications for the remaining conference try-outs. Findings and recommendations from the Regional Testing Conferences will be reported in the November 30, 1979 deliverable. Our findings will allow us to substantiate the claim of a workable alternative methodology for content/descriptive validity judgments at the test item and subtest level.

Dissemination

The results of this work are relevant for school personnel and curriculum developers. As districts and schools increasingly rely upon locally developed tests of objectives-based tests drawn from item banks, this method for making content validity judgments will assist in the development or selection of an instrument sensitive to instruction or curricula. Similarly, this systematic method of content validity will allow curriculum development agencies to maximize the sensitivity and descriptive power of their curriculum-embedded tests. Besides the obvious exposure through the Regional Testing Conferences to the audiences described above, further

dissemination plans include our roundtable paper of progress accepted for presentation at the 1979 Annual Meeting of the American Educational Research Association in San Francisco.

REFERENCES

Arbib, M. A. (Ed.). Algebraic theory of machines, languages and semigroups. New York, NY: Academic Press, 1968.

Carroll, J. B. Psychometric tests as cognitive tasks: A new "Structure of intellect." In L. B. Resnick (Ed.), The nature of intelligence. Hillsdale, NJ: Lawrence Earlbaum Associates, 1976.

Guion, R. M. Content validity--The source of my discontent. Applied Psychological Measurement, 1977, 1, 1-10.

Popham, W. J. Educational evaluation. Englewood Cliffs, NJ: Prentice-Hall, 1975.

Popham, W. J. Criterion-referenced measurement. Englewood Cliffs, NJ: Prentice-Hall, 1978.

Walker, C. B., Trithart, M. L., & Polin, L. G., Item-objective congruence in criterion-referenced tests: A method and some issues. Paper presented at the Annual Meeting of the California Educational Research Association, Burlingame, California, 1976.